

INFORMATION-GEOMETRIC EQUIVALENCE OF TRANSPORTATION POLYTOPES

MLADEN KOVAČEVIĆ, IVAN STANOJEVIĆ, AND VOJIN ŠENK

ABSTRACT. This paper deals with transportation polytopes in the probability simplex (that is, sets of categorical bivariate probability distributions with prescribed marginals). Information projections between such polytopes are studied, and a sufficient condition is described under which these mappings are homeomorphisms.

1. PRELIMINARIES

Let Γ_n denote the set of probability distributions with alphabet $\{1, \dots, n\}$:

$$(1.1) \quad \Gamma_n = \left\{ (p_i) \in \mathbb{R}^n : p_i \geq 0, \sum_i p_i = 1 \right\}.$$

The support of a probability distribution $P = (p_i)$ is denoted by $\text{supp}(P) = \{i : p_i > 0\}$, and its size by $|\text{supp}(P)|$. The support of a set \mathcal{P} of probability distributions is defined as $\text{supp}(\mathcal{P}) = \bigcup_{P \in \mathcal{P}} \text{supp}(P)$. If \mathcal{P} is convex, then there must exist $P \in \mathcal{P}$ with $\text{supp}(P) = \text{supp}(\mathcal{P})$. We will also write $P(i)$ for the masses of P .

Let $\mathcal{C}(P, Q)$ denote the set of all bivariate probability distributions with marginals $P \in \Gamma_n$ and $Q \in \Gamma_m$:

$$(1.2) \quad \mathcal{C}(P, Q) = \left\{ (s_{i,j}) \in \mathbb{R}^{n \times m} : s_{i,j} \geq 0, \sum_j s_{i,j} = p_i, \sum_i s_{i,j} = q_j \right\}.$$

Such sets are special cases of the so-called transportation polytopes, and have been studied extensively in probability, statistics, geometry, combinatorics, etc. (see, e.g., [2, 14]). In information-theoretic approaches to statistics, and in particular to the analysis of (multidimensional) contingency tables, a basic role is played by the so-called *information projections*, see [7] and the references therein. This motivates our study, presented in this note, of some formal properties of information projections (I-projections for short) over domains of the form $\mathcal{C}(P, Q)$. I-projections onto $\mathcal{C}(P, Q)$ also arise in binary hypothesis testing, see [13]. Further information-theoretic results (in a fairly different direction) regarding transportation polytopes can be found in [12].

Relative entropy (information divergence, Kullback-Leibler divergence) of the distribution P with respect to the distribution Q is defined by:

$$(1.3) \quad D(P||Q) = \sum_i p_i \log \frac{p_i}{q_i},$$

with the conventions $0 \log \frac{0}{q} = 0$ and $p \log \frac{p}{0} = \infty$ for every $q \geq 0$, $p > 0$, being understood. The functional D is nonnegative, equals zero if and only if $P = Q$, and is jointly convex in its arguments [6].

For a probability distribution S and a set of distributions \mathcal{T} , the I-projection [3, 4, 5, 15, 8] of S onto \mathcal{T} is defined as the unique minimizer (if it exists) of the functional $D(T||S)$ over all $T \in \mathcal{T}$. We shall study here I-projections as mappings between sets of the form $\mathcal{C}(P, Q)$. Namely, let $I_{\text{proj}} : \mathcal{C}(P_1, Q_1) \rightarrow \mathcal{C}(P_2, Q_2)$ be defined by:

$$(1.4) \quad I_{\text{proj}}(S) = \arg \inf_{T \in \mathcal{C}(P_2, Q_2)} D(T||S).$$

(Above and in the sequel we assume that $P_1, P_2 \in \Gamma_n$ and $Q_1, Q_2 \in \Gamma_m$.) The definition is slightly imprecise in that $I_{\text{proj}}(S)$ can be undefined for some $S \in \mathcal{C}(P_1, Q_1)$, i.e., the domain of I_{proj} can

Date: March 9, 2015.

2010 Mathematics Subject Classification. 94A17, 62B10, 62H17, 52B11, 52B12, 54C99.

Key words and phrases. Information projection, Kullback-Leibler divergence, transportation polytope, distributions with given marginals, contingency table, Fréchet-Hoeffding bounds.

in fact be a proper subset of $\mathcal{C}(P_1, Q_1)$. This is overlooked for notational simplicity. Another simplification is the omission of the dependence of the functional I_{proj} on P_i, Q_i ; this will not cause any ambiguities.

Note that $I_{\text{proj}}(S)$ is undefined only when $D(T||S) = \infty$ for all $T \in \mathcal{C}(P_2, Q_2)$. If $D(T||S) < \infty$ for some $T \in \mathcal{C}(P_2, Q_2)$, then existence of $I_{\text{proj}}(S)$ follows easily from the properties of $\mathcal{C}(P_2, Q_2)$ and the convexity of $D(\cdot||\cdot)$ [6]. Therefore, $I_{\text{proj}}(S)$ exists if and only if there exists $T \in \mathcal{C}(P_2, Q_2)$ with $\text{supp}(T) \subseteq \text{supp}(S)$. Furthermore, it is clear that the I-projection is defined for all $S \in \mathcal{C}(P_1, Q_1)$ if and only if it is defined for all vertices of $\mathcal{C}(P_1, Q_1)$.

2. GEOMETRIC EQUIVALENCE OF TRANSPORTATION POLYTOPES

The vertices of transportation polytopes are uniquely determined by their supports and can be characterized as follows: U is a vertex of $\mathcal{C}(P_1, Q_1)$ if and only if the associated bipartite graph G_U with “left” nodes $\{1, \dots, n\}$, “right” nodes $\{1, \dots, m\}$, and edges $\{(i, j) : U(i, j) > 0\}$, is a forest, i.e., contains no cycles [11]. In fact, every face of the polytope $\mathcal{C}(P_1, Q_1)$ is determined by its support [2]. Apart from identifying faces, the condition for two vertices being adjacent can also be expressed in terms of supports, as can many other geometric and combinatorial properties of transportation polytopes (see [9] and the references therein). This motivates the following definition.

Definition 2.1. We say that the polytopes $\mathcal{C}(P_1, Q_1)$ and $\mathcal{C}(P_2, Q_2)$ are *geometrically equivalent* if for every $S \in \mathcal{C}(P_1, Q_1)$ there exists $T \in \mathcal{C}(P_2, Q_2)$ with $\text{supp}(S) = \text{supp}(T)$, and vice versa. This is equivalent to saying that for every *vertex* $U \in \mathcal{C}(P_1, Q_1)$ there exists a *vertex* $V \in \mathcal{C}(P_2, Q_2)$ with $\text{supp}(U) = \text{supp}(V)$, and vice versa. \blacktriangle

Further justification of the term “geometrically equivalent”, in a certain information-geometric sense, is given in Theorem 3.3 below.

Example 2.2. To give an example of two geometrically equivalent transportation polytopes, consider some $\mathcal{C}(P_1, Q_1)$ that is generic (nondegenerate) [9], implying that the bipartite graphs defining its vertices are spanning trees, and assume that Q_1 has only two masses ($m = 2$). In this case for every vertex $U \in \mathcal{C}(P_1, Q_1)$, G_U has $n + 1$ edges and therefore necessarily contains edges $(i, 1)$ and $(i, 2)$ for some $i \in \{1, \dots, n\}$ (Fig. 1). Then it is not hard to see that $\mathcal{C}(P_1, Q_2)$ where

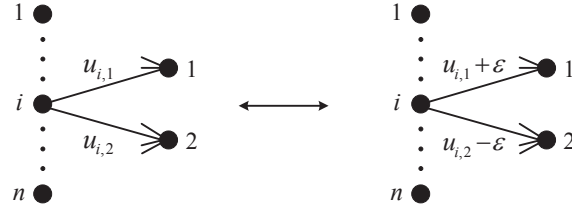


FIGURE 1. Graphs of the vertices of $\mathcal{C}(P_1, Q_1)$ and $\mathcal{C}(P_1, Q_2)$.

$Q_2(1) = Q_1(1) + \varepsilon$, $Q_2(2) = Q_1(2) - \varepsilon$, has vertices with identical supports as those of $\mathcal{C}(P_1, Q_1)$, for small enough ε . Thus, $\mathcal{C}(P_1, Q_1)$ and $\mathcal{C}(P_1, Q_2)$ are geometrically equivalent. \blacktriangle

The following claim is straightforward.

Proposition 2.3. If $\mathcal{C}(P_1, Q_1)$ and $\mathcal{C}(P_2, Q_2)$ are geometrically equivalent, then they are combinatorially equivalent, i.e., they have isomorphic face lattices. \blacksquare

3. I-PROJECTIONS BETWEEN TRANSPORTATION POLYTOPES

It is easy to see from the above discussion that I_{proj} maps the vertices of $\mathcal{C}(P_1, Q_1)$ to the vertices of $\mathcal{C}(P_2, Q_2)$. In the study of probability distributions with fixed marginals there are two particularly important vertices called Fréchet-Hoeffding (F-H for short) upper and lower bounds [14]. Both of them are uniquely determined by their supports, namely, $\overline{F} \in \mathcal{C}(P, Q)$ is the F-H upper bound for the family $\mathcal{C}(P, Q)$ if and only if its associated bipartite graph has no crossings (when the nodes $\{1, \dots, n\}$ on the left and $\{1, \dots, m\}$ on the right are drawn in increasing order, see Fig. 1), while $\underline{F} \in \mathcal{C}(P, Q)$ is the F-H lower bound if and only if its associated graph, after

reversing the order of the “right” nodes, has no crossings (in other words, the support of the lower bound for $\mathcal{C}(P, Q)$ is the same as that of the upper bound for $\mathcal{C}(P, \tilde{Q})$, where \tilde{Q} is the inverse permutation of Q , i.e., $\tilde{Q}(i) = Q(m + 1 - i)$). We then have:

Proposition 3.1. *Let $\bar{F}_i, \underline{E}_i \in \mathcal{C}(P_i, Q_i)$, $i \in \{1, 2\}$, be the F - H upper and lower bounds. If the I -projection of \bar{F}_1 (resp. \underline{E}_1) onto $\mathcal{C}(P_2, Q_2)$ exists, it is necessarily \bar{F}_2 (resp. \underline{E}_2).* ■

Another particular case that can be derived directly is that $I_{\text{proj}}(P_1 \times Q_1) = P_2 \times Q_2$, whenever $\text{supp}(P_2) \subseteq \text{supp}(P_1)$ and $\text{supp}(Q_2) \subseteq \text{supp}(Q_1)$. To prove this, it is by [7, Thm 3.2] enough to show that:

$$(3.1) \quad D(T \| P_1 \times Q_1) = D(P_2 \times Q_2 \| P_1 \times Q_1) + D(T \| P_2 \times Q_2)$$

for all $T \in \mathcal{C}(P_2, Q_2)$, which follows from:

$$(3.2) \quad \sum_{i,j} T(i,j) \log \frac{P_2(i)Q_2(j)}{P_1(i)Q_1(j)} = D(P_2 \| P_1) + D(Q_2 \| Q_1) = D(P_2 \times Q_2 \| P_1 \times Q_1).$$

We now restrict our attention to geometrically equivalent polytopes.

Proposition 3.2. *$\mathcal{C}(P_1, Q_1)$ and $\mathcal{C}(P_2, Q_2)$ are geometrically equivalent if and only if every vertex $U \in \mathcal{C}(P_1, Q_1)$ has an I -projection onto $\mathcal{C}(P_2, Q_2)$ and every vertex $V \in \mathcal{C}(P_2, Q_2)$ has an I -projection onto $\mathcal{C}(P_1, Q_1)$.*

Proof. The “only if” part is straightforward. For the “if” part, take some vertex $U \in \mathcal{C}(P_1, Q_1)$; let its I -projection onto $\mathcal{C}(P_2, Q_2)$ be U^* , and let the I -projection of U^* onto $\mathcal{C}(P_1, Q_1)$ be U' . We know that $\text{supp}(U') \subseteq \text{supp}(U^*) \subseteq \text{supp}(U)$, but in fact none of the inclusions can be strict because there can be no two vertices of a transportation polytope such that the support of one of them contains the support of the other. ■

The main result that we wish to report in this note is stated in the following theorem. It is a direct consequence of the propositions proved subsequently.

Theorem 3.3. *If $\mathcal{C}(P_1, Q_1)$ and $\mathcal{C}(P_2, Q_2)$ are geometrically equivalent, then they are homeomorphic under information projections.* ■

We first give a simple proof of continuity of information projections by using a well known identity obeyed by these functionals. See also [10] for a slightly different proof (obtained for the more general notion of f -projections). The assumed topology is the one induced by the ℓ_1 norm, and in what follows $P_n \rightarrow P$ means that $\|P_n - P\|_1 \equiv \sum_i |P_n(i) - P(i)| \rightarrow 0$.

Proposition 3.4. *I_{proj} is continuous in its domain.*

Proof. Let $S_n, S \in \mathcal{C}(P_1, Q_1)$ with $S_n \rightarrow S$. Let $S^* = I_{\text{proj}}(S)$, $S_n^* = I_{\text{proj}}(S_n)$; we need to show that $S_n^* \rightarrow S^*$. Since $\mathcal{C}(P_2, Q_2)$ is compact, S_n^* must have a convergent subsequence $S_{k_n}^*$ (k_n is an increasing function in n). Suppose that $S_{k_n}^* \rightarrow R$ for some $R \in \mathcal{C}(P_2, Q_2)$. The set of all distributions $T \in \mathcal{C}(P_2, Q_2)$ with $\text{supp}(T) \subseteq \text{supp}(S_{k_n})$ is a linear family¹ [7], and therefore the following identity holds [7, Thm 3.2]:

$$(3.3) \quad D(T \| S_{k_n}) = D(S_{k_n}^* \| S_{k_n}) + D(T \| S_{k_n}^*)$$

for all $T \in \mathcal{C}(P_2, Q_2)$ with $\text{supp}(T) \subseteq \text{supp}(S_{k_n})$. Taking the limit when $n \rightarrow \infty$ and using the fact that $D(\cdot \| \cdot)$ is continuous in its second argument (in the finite alphabet case), we obtain:

$$(3.4) \quad D(T \| S) = \lim_{n \rightarrow \infty} D(S_{k_n}^* \| S_{k_n}) + D(T \| R).$$

Evaluating (3.4) at $T = R$ we conclude that $\lim_{n \rightarrow \infty} D(S_{k_n}^* \| S_{k_n}) = D(R \| S)$. Substituting this back into (3.4) and evaluating at $T = S^*$ we get:

$$(3.5) \quad D(S^* \| S) = D(R \| S) + D(S^* \| R),$$

wherefrom $D(S^* \| S) \geq D(R \| S)$. But since S^* is by assumption the unique minimizer of $D(\cdot \| S)$ over $\mathcal{C}(P_2, Q_2)$, we must have $R = S^*$. ■

¹ A linear family of (two-dimensional) probability distributions is a set of the form $\{T : \sum_{i,j} T(i,j)f_k(i,j) = \alpha_k\}$, where f_k , $1 \leq k \leq K$, are real functions defined on the alphabet of the distributions T , and α_k are real numbers.

Proposition 3.5. *Let $\mathcal{C}(P_1, Q_1)$ and $\mathcal{C}(P_2, Q_2)$ be geometrically equivalent. Then I_{proj} is a bijection².*

Proof. 1.) I_{proj} is injective (one-to-one). Observe that every distribution $S \in \mathcal{C}(P_1, Q_1)$ maps to a distribution with the same support, i.e., $\text{supp}(I_{\text{proj}}(S)) = \text{supp}(S)$; this follows from [7, Thm 3.1] (that such a distribution exists follows from geometric equivalence of $\mathcal{C}(P_1, Q_1)$ and $\mathcal{C}(P_2, Q_2)$). We conclude that a vertex $V \in \mathcal{C}(P_1, Q_1)$ maps to the corresponding vertex $V^* \in \mathcal{C}(P_2, Q_2)$ with $\text{supp}(V^*) = \text{supp}(V)$, and no other distribution from $\mathcal{C}(P_1, Q_1)$ can map to V^* because vertices are uniquely determined by their supports. Assume now, for the sake of contradiction, that $I_{\text{proj}}(S_1) = I_{\text{proj}}(S_2) = S^*$, where $S_1, S_2 \in \mathcal{C}(P_1, Q_1)$ are not vertices. As commented above, we necessarily have $\text{supp}(S_1) = \text{supp}(S_2) = \text{supp}(S^*)$. Furthermore, by [7, Thm 3.2] we have:

$$(3.6) \quad \begin{aligned} D(T||S_1) &= D(S^*||S_1) + D(T||S^*) \\ D(T||S_2) &= D(S^*||S_2) + D(T||S^*) \end{aligned}$$

and by subtracting these equations we get:

$$(3.7) \quad D(T||S_1) - D(T||S_2) = D(S^*||S_1) - D(S^*||S_2)$$

for all $T \in \mathcal{C}(P_2, Q_2)$ with $\text{supp}(T) \subseteq \text{supp}(S^*)$. By writing out all terms of (3.7) we obtain:

$$(3.8) \quad \sum_{i,j} T(i,j) \log \frac{S_2(i,j)}{S_1(i,j)} = \sum_{i,j} S^*(i,j) \log \frac{S_2(i,j)}{S_1(i,j)}.$$

Define $\epsilon(i,j) = S_2(i,j) - S_1(i,j)$. We can evaluate (3.8) at $T = S^* + \delta\epsilon$ for some small enough constant $\delta > 0$, because $\sum_i \epsilon(i,j) = \sum_j \epsilon(i,j) = 0$ and $\text{supp}(S^*) = \text{supp}(S_1) = \text{supp}(S_2)$, which ensures that $S^* + \delta\epsilon \in \mathcal{C}(P_2, Q_2)$ and $\text{supp}(T) \subseteq \text{supp}(S^*)$. This gives:

$$(3.9) \quad \sum_{i,j} \epsilon(i,j) \log \frac{S_2(i,j)}{S_1(i,j)} = 0.$$

But $\epsilon(i,j)$ and $\log \frac{S_2(i,j)}{S_1(i,j)}$ always have the same sign, which means that the left-hand side of (3.9) is strictly positive and cannot equal zero, a contradiction.

2.) I_{proj} is surjective (onto). Let $\mathcal{F}_k \subseteq \mathcal{C}(P_1, Q_1)$ be a k -dimensional face of $\mathcal{C}(P_1, Q_1)$, $k \leq (n-1)(m-1)$, determined uniquely by its support $\text{supp}(\mathcal{F}_k)$, namely, $\mathcal{F}_k = \{S \in \mathcal{C}(P_1, Q_1) : \text{supp}(S) \subseteq \text{supp}(\mathcal{F}_k)\}$. We can regard \mathcal{F}_k as a convex and compact subset of its affine hull, denoted $\text{aff}(\mathcal{F}_k)$. When regarded this way, the interior of \mathcal{F}_k is nonempty and consists of distributions with full support, namely, $\text{int}(\mathcal{F}_k) = \{S \in \mathcal{F}_k : \text{supp}(S) = \text{supp}(\mathcal{F}_k)\}$. The boundary of \mathcal{F}_k , denoted $\partial\mathcal{F}_k$, is the union of the proper faces of \mathcal{F}_k . Distributions in $\partial\mathcal{F}_k$ have supports strictly contained in $\text{supp}(\mathcal{F}_k)$. Now, let \mathcal{F}_k^* be the corresponding face of $\mathcal{C}(P_2, Q_2)$ with $\text{supp}(\mathcal{F}_k^*) = \text{supp}(\mathcal{F}_k)$. We know that I_{proj} maps distributions from \mathcal{F}_k to distributions from \mathcal{F}_k^* ($I_{\text{proj}}(\mathcal{F}_k) \subseteq \mathcal{F}_k^*$) because, for $S \in \mathcal{F}_k$, $D(\cdot||S)$ is finite only over \mathcal{F}_k^* . We will show that in fact $I_{\text{proj}}(\mathcal{F}_k) = \mathcal{F}_k^*$, i.e., that I_{proj} is surjective over \mathcal{F}_k , which will establish the desired claim. The proof is by induction on the dimension of the faces (k). We first observe, again by analyzing supports, that $I_{\text{proj}}(\text{int}(\mathcal{F}_k)) \subseteq \text{int}(\mathcal{F}_k^*)$, and $I_{\text{proj}}(\partial\mathcal{F}_k) \subseteq \partial\mathcal{F}_k^*$ (in fact, the image of every proper face of \mathcal{F}_k is contained in the corresponding face of \mathcal{F}_k^* having the same support). We can now start the induction. Assume that I_{proj} is surjective over every face of $\mathcal{C}(P_1, Q_1)$ of dimension $< k$ (we know that it is surjective over zero-dimensional faces, i.e., vertices, and so the induction is justified). Therefore, the assumption is that $I_{\text{proj}}(\partial\mathcal{F}_k) = \partial\mathcal{F}_k^*$, and we need to show that also $I_{\text{proj}}(\text{int}(\mathcal{F}_k)) = \text{int}(\mathcal{F}_k^*)$. We will use the following simple claim.

Claim 1. *Let A and B be open sets (in arbitrary topological space) with $A \subseteq B$, and B connected. If A and B have the same boundaries ($\partial A = \partial B$) then they are equal.*

Proof: Assume that $A \neq B$, and let $x \in B \setminus A$. There must exist a neighborhood of x , denoted $V(x)$, such that $V(x) \subseteq B \setminus A$ for otherwise we would have that $x \in \partial A = \partial B$ which is impossible since B is open and cannot contain its boundary points. This proves that $B \setminus A$ is open and hence B is a union of two disjoint open sets (A and $B \setminus A$). This is a contradiction because B is connected. \blacktriangle

² Note that this follows from a stronger statement given in Proposition 3.6, but we also give here a direct proof that we believe is interesting in its own right.

We know that $I_{\text{proj}}(\text{int}(\mathcal{F}_k)) \subseteq \text{int}(\mathcal{F}_k^*)$, and that $\text{int}(\mathcal{F}_k^*)$ is open (in $\text{aff}(\mathcal{F}_k^*)$) and connected. Hence, to prove that $I_{\text{proj}}(\text{int}(\mathcal{F}_k)) = \text{int}(\mathcal{F}_k^*)$ (by using Claim 1), we need to show that $I_{\text{proj}}(\text{int}(\mathcal{F}_k))$ is open, and that $\partial I_{\text{proj}}(\text{int}(\mathcal{F}_k)) = \partial \text{int}(\mathcal{F}_k^*) \equiv \partial \mathcal{F}_k^*$. Since I_{proj} is an injective and continuous function from a compact to a metric space, it is a homeomorphism onto its image [1, Thm 7.8, Ch I]. In particular, it is both open and closed. Therefore, $I_{\text{proj}}(\text{int}(\mathcal{F}_k))$ is indeed open in $\text{aff}(\mathcal{F}_k^*)$. Furthermore, $I_{\text{proj}}(\mathcal{F}_k) = I_{\text{proj}}(\text{int}(\mathcal{F}_k)) \cup \partial \mathcal{F}_k^*$ is closed in $\text{aff}(\mathcal{F}_k^*)$, which implies that the boundary of $I_{\text{proj}}(\text{int}(\mathcal{F}_k))$ is contained in $\partial \mathcal{F}_k^*$. But in fact it must be equal to $\partial \mathcal{F}_k^*$ because any $T^* \in \partial \mathcal{F}_k^*$ is a limit point of $I_{\text{proj}}(\text{int}(\mathcal{F}_k))$. Namely, T^* must be the image of some $T \in \partial \mathcal{F}_k$ by the induction hypothesis, and if $T_n \rightarrow T$, $T_n \in \text{int}(\mathcal{F}_k)$, then $I_{\text{proj}}(T_n) \rightarrow T^*$ by continuity. The proof is complete. ■

In the above proof we used the fact that the inverse of the I-projection from $\mathcal{C}(P_1, Q_1)$ to $\mathcal{C}(P_2, Q_2)$ is continuous. The following proposition precisely identifies this inverse. The statement is somewhat counterintuitive due to the asymmetry of the functional $D(\cdot|\cdot)$.

Proposition 3.6. *Let $\mathcal{C}(P_1, Q_1)$ and $\mathcal{C}(P_2, Q_2)$ be geometrically equivalent. Then the inverse of the I-projection from $\mathcal{C}(P_1, Q_1)$ to $\mathcal{C}(P_2, Q_2)$ is the I-projection from $\mathcal{C}(P_2, Q_2)$ to $\mathcal{C}(P_1, Q_1)$.*

Proof. The linear families $\mathcal{C}(P_1, Q_1)$ and $\mathcal{C}(P_2, Q_2)$ are translates³ of each other in the sense of [7]. Let $S \in \mathcal{C}(P_1, Q_1)$ and let S^* be its I-projection onto $\mathcal{C}(P_2, Q_2)$. By [7, Lemma 4.2], the I-projections of S and S^* onto $\mathcal{C}(P_1, Q_1)$ must be identical, and this is trivially S . (Apart from being translates of each other, the additional condition of [7, Lemma 4.2] dealing with supports is also satisfied due to geometric equivalence of $\mathcal{C}(P_1, Q_1)$ and $\mathcal{C}(P_2, Q_2)$.) ■

We conclude the paper by illustrating that the converse of Theorem 3.3 does not hold. The following example exhibits two transportation polytopes that are not geometrically equivalent, but are homeomorphic under information projection.

Example 3.7. Let $P_1 = (1/2, 1/2)$, $Q_1 = (1/3, 2/3)$, and $P_2 = Q_2 = (1/2, 1/2)$. Both $\mathcal{C}(P_1, Q_1)$ and $\mathcal{C}(P_2, Q_2)$ are one-dimensional polytopes, but clearly not geometrically equivalent because their vertices are:

$$(3.10) \quad U_1 = \begin{pmatrix} 1/3 & 1/6 \\ 0 & 1/2 \end{pmatrix}, \quad U_2 = \begin{pmatrix} 0 & 1/2 \\ 1/3 & 1/6 \end{pmatrix}$$

for $\mathcal{C}(P_1, Q_1)$ and

$$(3.11) \quad V_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}, \quad V_2 = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}$$

for $\mathcal{C}(P_2, Q_2)$. Let I_{proj} denote the I-projection from $\mathcal{C}(P_1, Q_1)$ to $\mathcal{C}(P_2, Q_2)$, as before. I_{proj} is continuous by Proposition 3.4. By using [7, Lemma 4.2] in the same way as in Proposition 3.6, one can show that it is bijective over the interior of $\mathcal{C}(P_1, Q_1)$ (which consists of distributions from $\mathcal{C}(P_1, Q_1)$ having full support), and that its inverse over this domain is precisely the I-projection from $\mathcal{C}(P_2, Q_2)$ to $\mathcal{C}(P_1, Q_1)$. Since $I_{\text{proj}}(U_i) = V_i$, $i \in \{1, 2\}$, I_{proj} is bijective over the entire $\mathcal{C}(P_1, Q_1)$, and hence it is a homeomorphism. Its inverse is guaranteed to be continuous by [1, Thm 7.8, Ch I], but note that this inverse is *not* the I-projection from $\mathcal{C}(P_2, Q_2)$ to $\mathcal{C}(P_1, Q_1)$ because the I-projection of V_i onto $\mathcal{C}(P_1, Q_1)$ is undefined. ▲

ACKNOWLEDGMENT

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (grants TR32040 and III44003).

REFERENCES

1. G. E. Bredon, *Topology and Geometry*, Springer-Verlag, 1993.
2. R. A. Brualdi, *Combinatorial Matrix Classes*, Cambridge University Press, 2006.
3. N. N. Chentsov, "A Nonsymmetric Distance Between Probability Distributions, Entropy and the Pythagorean Theorem," *Math. Notes*, vol. 4, no. 3, pp. 686-691, Sept. 1968.
4. N. N. Chentsov, *Statistical Decision Rules and Optimal Inference*, (in Russian). Providence, RI: Translations of Mathematical Monographs, Amer. Math. Soc., 1982. Original publication: Moscow, U.S.S.R.: Nauka, 1972.

³ Linear families are translates of each other if they are defined by the same functions f_k but different numbers α_k .

5. I. Csiszár, "I-Divergence Geometry of Probability Distributions and Minimization Problems," *Ann. Probab.*, vol. 3, no. 1, pp. 146–158, 1975.
6. I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, Inc., 1981.
7. I. Csiszár and P. Shields, "Information Theory and Statistics: A Tutorial," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 4, pp. 417–528, Dec. 2004.
8. I. Csiszár and F. Matúš, "Information Projections Revisited," *IEEE Trans. Inform. Theory*, vol. 49, no. 6, pp. 1474–1490, June 2003.
9. J. A. De Loera and E. D. Kim, "Combinatorics and Geometry of Transportation Polytopes: An Update," In *Discrete Geometry and Algebraic Combinatorics*, vol. 625 of *Contemporary Mathematics*, pp. 37–76, American Mathematical Society, Providence, RI, 2014.
10. C. Gietl and F. P. Ruffel, "Continuity of f-projections on Discrete Spaces," in *Geometric Science of Information, Lecture Notes in Computer Science*, vol. 8085, pp. 519–524, 2013.
11. V. Klee and C. Witzgall, "Facets and vertices of transportation polytopes," In: *Mathematics of the Decision Sciences, Part I* (Stanford, CA, 1967), 257–282, AMS, Providence, RI, 1968.
12. M. Kovačević, I. Stanojević, and V. Šenk, "On the Entropy of Couplings," *Inform. and Comput.*, vol. 242, pp. 369–382, June 2015.
13. Y. Polyanskiy, "Hypothesis Testing via a Comparator," in *Proc. 2012 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2206–2210, Cambridge, MA, July 2012.
14. L. Rüschendorf, B. Schweizer, and M. D. Taylor (Editors), *Distributions with Fixed Marginals and Related Topics*, Lecture Notes - Monograph Series, Institute of Mathematical Statistics, 1996.
15. F. Topsøe, "Information Theoretical Optimization Techniques," *Kybernetika*, vol. 15, no. 1, pp. 8–27, 1979.

DEPARTMENT OF ELECTRICAL ENGINEERING, FACULTY OF TECHNICAL SCIENCES, UNIVERSITY OF NOVI SAD, TRG
DOSITEJA OBRADOVIĆA 6, 21000 NOVI SAD, SERBIA

E-mail address: `kmladen@uns.ac.rs` (corresponding author), `cet.ivan@uns.ac.rs`, `vojin_senk@uns.ac.rs`